AD-A119 715 RENSSELAER POLYTECHNIC INST TROY NY F/G 12/1

SELF-CRITICAL, AND ROBUST, PROCEDURES FOR THE ANALYSIS OF MULTI--ETC(U)

JUN 82 A S PAULSON, N J DELANEY DAAG29-81-K-0110

UNCLASSIFIED A-4 ARO-18072.4-MA NL

END

ARO 18072.4-MA

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AD A119715

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: A-4	2. GOVT ACCESSION NO.: AD-A119 715
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): SELF-CRITICAL, AND ROBUST, PROCEDURES FOR THE ANALYSIS OF MULTIVARIATE NORMAL DATA	5. TYPE OF REPORT & PERIOD COVERED: Interim Technical Report
	6. PERFORMING ORG. REPORT NUMBER: A-4
7. AUTHOR(s): A. S. Paulson, N. J. Delaney	8. CONTRACT OR GRANT NUMBER(s): DAA G29-81-K-0110
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Rensselaer Polytechnic Institute, Troy, New York 12181	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Approved for public release; distribution unlimited.	12. REPORT DATE: 1 June 1982
	13. NUMBER OF PAGES: 40
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Department of the Navy, Office of Naval Research, 715 Broadway (5th Floor), New York, New York 10003	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

U. S. Army Research Office
Post Office Box 12211
Research Triangle Park, NC 27709

This document has been approved for public release and sale; its distribution is unlimited.

DTIC ELECTE SEP 28 1982 S D H

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

THE VIEW, OPINIONS, AND/OR FINDINGS CONTAINED IN THIS REPORT ARE THOSE OF THE AUTHOR(S) AND SHOULD NOT BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, POLICY, OR DECISION, UNLESS SO DESIGNATED BY OTHER DOCUMENTATION.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

self-critical, multivariate normal, influence functions, generalized likelihood, objective function, clustering

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A procedure for performing sensitivity analysis of data vis-a-vis the multivariate error model, with or without structural variables, is proposed. It is based on an estimation procedure, a generalization of maximum likelihood, which is indexed on a user-specified value c. The sensitivity analysis proceeds in the following way: the parameters of a set of data presumed to follow a multivariate normal error model are estimated via maximum likelihood (c=0). The parameters are next estimated for an index $c_1>0$.

DTIC FILE COPY

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

82 09 28 002

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (cont'd)

The sensitivity in parameter estimates is noted. The parameters are next estimated for a user-specified index $c_2>c_1>0$ and the sensitivity of the parameter estimates to the change in c is again noted. This process may have one or more values of c>0. If the parameter estimates are sensitive functions of c, the model and the data are not mutually consistent and both require further detailed study. The items of data which are most contributory to this sensitivity are identified by an examination of observational weights which are a byproduct of the analysis. If the parameters and observational weights are non-sensitive to changes in the index c, then one can generally be confident of the internal consistency of the data and the error model. Fixed values of the index c provide robust estimation procedures for model parameters. Asymptotic relative efficiencies and influence functions are provided for fixed values of c. The results of a small Monte Carlo study suggest that the asymptotic properties of the estimators are rapidly attained. Several illustrations are given.

SELF-CRITICAL, AND ROBUST, PROCEDURES FOR THE ANALYSIS OF MULTIVARIATE NORMAL DATA

by

A.S. Paulson *
Rensselaer Polytechnic Institute

and

N.J. Delaney
Union College

*Research supported in part by U.S. Army Research Office under contract DAA G29-81-K-0110.

Key Words: self-critical, multivariate normal, influence functions, generalized likelihood, objective function, clustering.

ABSTRACT

A procedure for performing sensitivity analyses of data vis-à-vis the multivariate error model, with or without structural variables, is proposed. It is based on an estimation procedure, a generalization of maximum likelihood, which is indexed on a user-specified value c. The sensitivity analysis proceeds in the following way: the parameters of a set of data presumed to follow a multivariate normal error model are estimated via maximum likelihood (c=0). The parameters are next estimated for an index $c_1 > 0$. The sensitivity in parameter estimates is noted. The parameters are next estimated for a user-specified index $c_2 > c_1 > 0$ and the sensitivity of the parameter estimates to the change in c is again noted. This process may have one or more values of $c > 0$. If the parameter estimates are sensitive functions of c, the model and the data are not mutually consistent and both require further detailed study. The items of data which are most contributory to this sensitivity are identified by an examination of observational weights which are a byproduct of the analysis. If the parameters and observational weights are non-sensitive to changes in the index c, then one can generally be confident of the internal consistency of the data and the error model. Fixed values of the index c provide robust estimation procedures for model parameters. Asymptotic relative efficiencies and influence functions are provided for fixed values of c. The results of a small Monte Carlo study suggest that the asymptotic properties of the estimators are rapidly attained. Several illustrations are given.

1. Introduction

A number of robust procedures for the analysis of p-dimensional Gaussian (normal) data have been proposed by Gnanadesikan and Kettenring (1972), Devlin, Gnanadesikan, and Kettenring (1975), Maronna (1976), Huber (1977), Mosteller and Tukey (1977), Devlin, Gnanadesikan, and Kettenring (1981). A number of numerical and statistical characteristics of these procedures have been evaluated in Devlin et al. (1981). Summaries and further discussion of multivariate robust procedures are given in Gnanadesikan (1977), Barnett and Lewis (1978), and Huber (1981). Despite the impressive range of work in the problem of robustness and allied topics, there is still a need for a unified approach to the subject, preferably one based on an objective function.

We propose a procedure for the analysis of multivariate Gaussian data which is based on a generalization of maximum likelihood which has recently been considered in the univariate context by Paulson, Presser, and Nicklin (1982). The estimation procedure depends on a user-specified index c which may be varied to determine (a), the response of the parameter estimates to variation in this index, and (b), the response of observational weights associated with each observation to variation in this index. The response surface generated permits a sensitivity analysis which is useful in assessing the mutual consistency of the model and the data considered as a single entity. If the parameter estimates and final observational weights are not sensitive to variation in the user-specified index, then it is highly likely that the data and the model are internally consistent. If the parameter estimates and final weights are sensitive to variation in the user-specified index, then it is likely that the data and the model are not internally consistent for any of a variety of reasons. Thus the procedure we are proposing can be quite useful in the identification of potential outliers.

Our procedure produces robust estimators of the mean vector and the covariance matrix by simply taking a fixed value of the user-specified index c. Asymptotic properties are given for several values of the user-specified constant and for several values of the dimension p. The procedure is easy to use and computationally attractive. Several examples and illustrations are provided.

2. The Simultaneous Estimation Procedure

We suppose that $\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n$ constitutes a random sample from, tentatively, a multivariate Gaussian density

$$f(\underset{\sim}{x}|\underset{\sim}{\mu}, \underset{\sim}{D}) = |2\pi\underset{\sim}{D}|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}(\underset{\sim}{x}-\underset{\sim}{\mu})^T \underset{\sim}{D}^{-1}(\underset{\sim}{x}-\underset{\sim}{\mu})\} \tag{2.1}$$

where $\underset{\sim}{\mu}$ is a p×1 vector of location parameters, and $\underset{\sim}{D} = (d_{jk})$ is a p×p positive definite matrix of covariances. The estimators for $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$, which we shall term self-critical, are determined from maximization of

$$\ell_c = \frac{1}{c} \sum_{i=1}^{n} \left\{ \frac{f^c(\underset{\sim}{x}_i|\underset{\sim}{\mu},\underset{\sim}{D})}{[Q(\underset{\sim}{\mu},\underset{\sim}{D};c)]^{c/(1+c)}} - 1 \right\} , \tag{2.2}$$

where

$$Q(\underset{\sim}{\mu},\underset{\sim}{D};c) = \int_{R_p} f^{1+c}(\underset{\sim}{x}|\underset{\sim}{\mu},\underset{\sim}{D})d\underset{\sim}{x}$$

$$= \{(1+c)^p (2\pi)^{cp}|\underset{\sim}{D}|^c\}^{-\frac{1}{2}} . \tag{2.3}$$

It is easily shown that

$$\lim_{c \to 0} \ell_c = \ell_0 = \sum_{i=1}^{n} \log f(\underset{\sim}{x}_i | \underset{\sim}{\mu}, \underset{\sim}{D}),$$

the likelihood function. Upon differentiation of (2.2) with respect to $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$, and setting the resulting expressions to $\underset{\sim}{0}$, we find that the estimators for $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$ are determined from the simultaneous zeros of

$$\sum_{i=1}^{n} f_i^{\,c} \{(1+c) \frac{\partial \log f_i}{\partial \underset{\sim}{\mu}} - \frac{\partial \log Q}{\partial \underset{\sim}{\mu}}\} = \underset{\sim}{0} \quad , \tag{2.4}$$

$$\sum_{i=1}^{n} f_i^{\,c} \{(1+c) \frac{\partial \log f_i}{\partial \underset{\sim}{D}} \quad \frac{\partial \log Q}{\partial \underset{\sim}{D}}\} = \underset{\sim}{0} \quad , \tag{2.5}$$

where we have suppressed the arguments of the density and Q and written f_i for $f(x_i | \underset{\sim}{\mu}, \underset{\sim}{D})$ and Q for $Q(\underset{\sim}{\mu}, \underset{\sim}{D}; c)$. Equations (2.4) and (2.5) may be rearranged so that the estimators for $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$, say $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{D}}$, satisfy the implicit equations

$$\underset{\sim}{\mu} = \frac{\sum_{i=1}^{n} \underset{\sim}{x}_i \, v_{ic}}{\sum_{i=1}^{n} v_{ic}} \quad , \tag{2.6}$$

$$\underset{\sim}{D} = (1+c) \frac{\sum_{i=1}^{n} (\underset{\sim}{x}_i - \underset{\sim}{\mu})(\underset{\sim}{x}_i - \underset{\sim}{\mu})^T v_{ic}}{\sum_{i=1}^{n} v_{ic}} \quad , \tag{2.7}$$

where

$$v_{ic} = \exp(-\frac{c}{2} (\underset{\sim}{x}_i - \underset{\sim}{\mu})^T \underset{\sim}{D}^{-1} (\underset{\sim}{x}_i - \underset{\sim}{\mu})) \; . \tag{2.8}$$

The weighting in (2.6) and (2.7) reflects a self-weighting, namely by the density itself. Thus the assumed density and the data interact in producing parameter estimators, and for this reason we have called the resulting estimators self-critical.

The system (2.6)-(2.7) defines a family of estimators indexed on the user-specified constant c. We will presently address the question of how c should be chosen for practical applications. It is recommended that the estimators be computed numerically by a fixed point algorithm with initial trial solutions the usual maximum likelihood estimators of $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$. Since the function ℓ_0 is concave (Rao, 1965, pp. 447-449), ℓ_c will also be concave for at least a neighborhood of $c = 0$ and will, for this neighborhood, always have a unique solution. The system (2.6)-(2.8) need not always have a unique solution in general but we have not encountered any difficulties with the estimation procedure in a wide variety of experiences with real and simulated data so long as the trial solution for $c > 0$ was the maximum likelihood ($c=0$) estimators and c was not too far from zero. As c moves further and further away from zero, the procedure may break down numerically or provide multiple solutions.

It is easily verified that the expectations of the left-hand sides of (2.4) and (2.5) are the zero vector and matrix respectively. In fact, the procedure was developed to guarantee just this; see Paulson et al. (1982) for further discussion.

3. Some Examples

A few examples at this point will be useful in showing the potential of the family of estimators (2.6)-(2.8) in the analysis of multivariate data.

Example 3.1

Consider the bivariate data of Figure 1 (Anderson (1958)). The central cloud of points represents the original data to which a single outlier has been added and rotated through 360° in 9° increments starting at the point labeled 1. Table 1 presents the results of the traditional maximum likelihood analysis applied to the original data plus the single outlier contrasted with the self-critical analysis with c = .3. The effect of the single outlier on the estimate of the means μ_1 and μ_2 is slight for the c = .3 case. The effect of the single outlier on the estimates of the variances σ_1^2 and σ_2^2 is substantial for c = 0 but is less than 23% for c = .3 regardless of where the outlier is placed. The effect on the estimate of correlation for c = 0 is dramatic, the estimate varying from .80 to .27. However, for c = .30 the estimate of correlation varies only from .746 and .800. In this case, graphical methods make it easy to see what is happening, but when the dimensionality p increases, it becomes increasingly difficult to isolate problems, or more importantly, the elements of the data which just might lead to the most interesting structure in the data. The automatic weighting of the observations via the self-critical procedure provides a means at higher dimensions for indicating unusual observations or structure. The weight $\hat{v}_{j.3}$ for the outlying observations is the lowest among all the weights.

Example 3.2

The extent to which the family of estimators (2.6)-(2.8) can be useful in identifying structure is now considered. Each of the 25 3-vector observations obtained by Billmeyer and Rich (1978) from color matching on

Figure 1

Position of Outlier Under Rotation Through 360°

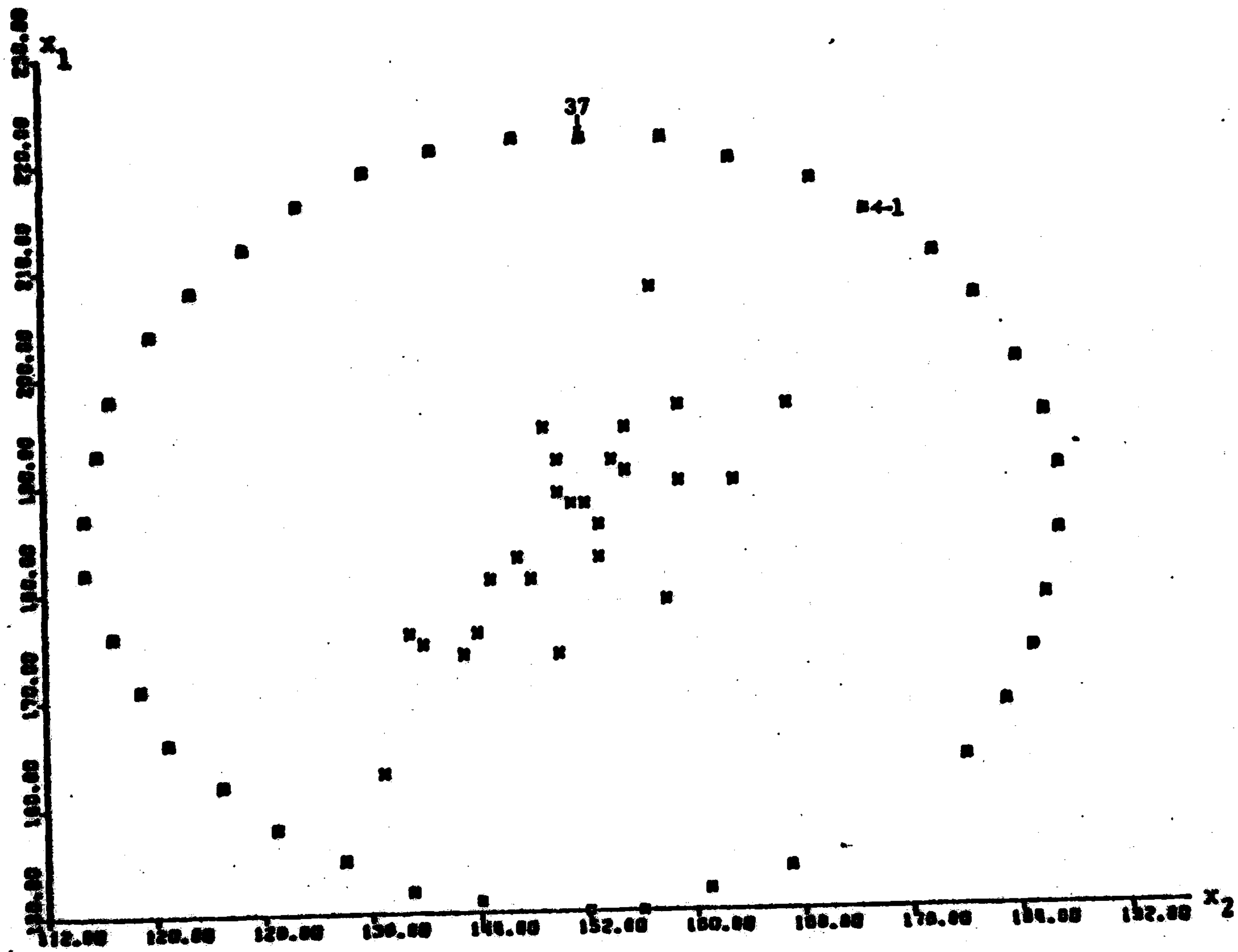


TABLE 1

Percent Change of Estimators for Location, Dispersion and Orientation Under Rotation of Outlier Through 9 Degree Increments

Parameter	$c = 0$	$c = .3$
μ_1	±1%	±0.3%
μ_2	±1%	±0.3%
σ_1^2	-4% to +50%	0% to +23%
σ_2^2	-4% to +88%	0% to +21%
ρ	-63% to +9%	-2% to +6%

a MacAdam Colorimeter are shown in Table 2. The three variables are the percentages of red, blue, and green in the final adjustment by the observer in his attempt to match a given color. Analysis of similar data, obtained from 22 sessions with a single observer, led Brown et al. (1956) to conclude that the data were trivariate normal. [The data are all ostensibly from a single trivariate population.] However, normal probability plots of each variable would show that observations 16, 19, 20, 21, 22, and 25 seem the most unusual in one or more dimensions and an Andrews-type plot corroborates this.

We applied equations (2.6)-(2.8) to these data, starting first at $c = 0$ and then progressively increasing c through .1, .3, to .5. Several observations are immediately highlighted by a dramatic change in the final weights $\hat{w}_{jc}$. Several components of the covariance matrix also show marked sensitivity to the variation in c. As c increases to .5, 12 of the weights are reduced to low values. The results for $c = .5$ are also presented in Table 2. The low weights $\hat{w}_{j,.5} = \hat{v}_{j,.5}/\Sigma\hat{v}_{j,.5}$ in Table 2 pinpoint observations 14-25 for further study. This analysis led to careful reexamination of the original data cards which showed that the first 13 observations were those of a single observer, A say, while the last 12 were those of 12 different observers.

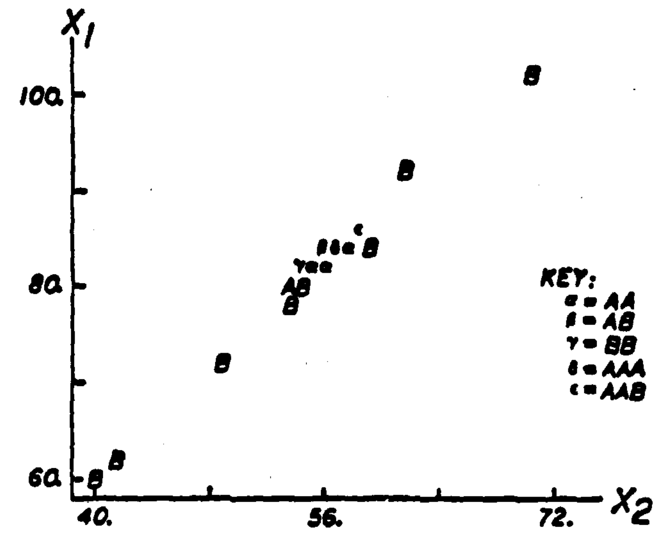
Figures 2a, b, and c show all two-dimensional scatter plots of this sample with observer A's observations labeled by A's and all other observations labeled by B's. Greek letters represent various combinations of observations which are too close to differentiate in a plot of this scale. For example, ε represents 2 A observations and 1 B observation.

TABLE 2

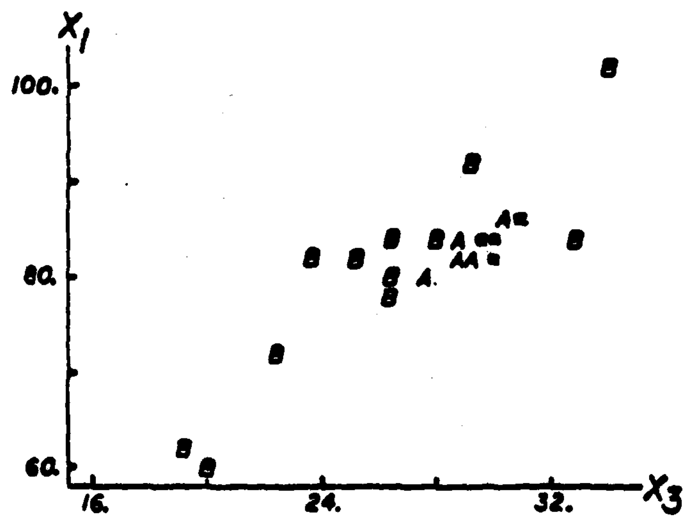
Sensitivity Analysis of Trivariate Data

				c=.125	c=.3	c=.5
Point No.	X_1	X_2	X_3	$\hat{w}_{j,.125}$	$\hat{w}_{j,.3}$	$\hat{w}_{j,.5}$
1	86.07	58.27	30.72	.051	.064	0.102
2	83.45	56.50	29.45	.053	.077	0.115
3	82.07	55.43	28.71	.053	.075	0.097
4	83.58	56.32	28.85	.049	.060	0.043
5	82.95	56.21	30.12	.050	.065	0.053
6	83.62	56.38	29.51	.048	.059	0.086
7	86.25	58.39	30.92	.050	.063	0.092
8	84.79	57.62	30.15	.047	.046	0.050
9	82.87	55.99	29.87	.050	.065	0.064
10	79.74	53.89	27.70	.053	.055	0.036
11	81.84	55.26	29.10	.051	.065	0.091
12	84.10	57.08	30.08	.049	.061	0.085
13	86.24	58.34	30.51	.051	.061	0.087
14	80.55	54.32	26.23	.051	.057	0.2(-4)
15	81.08	54.54	25.32	.046	.036	0.2(-8)
16	101.2	70.38	33.97	.1(-4)	.1(-18)	0.6(-41)
17	84.92	57.75	28.08	.033	.011	0.3(-6)
18	83.78	56.52	26.49	.044	.030	0.1(-8)
19	60.54	40.34	20.06	.020	.7(-8)	0.8(-14)
20	62.33	41.38	19.11	.024	.2(-6)	0.2(-13)
21	72.09	48.46	22.50	.042	.004	0.2(-8)
22	83.38	58.98	32.66	.5(-7)	.8(-23)	0.2(-39)
23	81.45	54.53	23.78	.030	.007	0.6(-18)
24	78.89	53.37	26.21	.048	.039	0.2(-3)
25	92.50	61.76	29.02	.007	.3(-5)	0.1(-16)

Figure 2



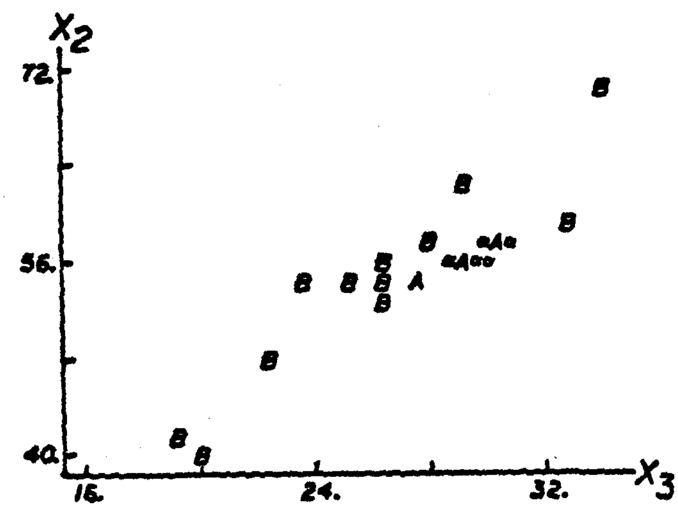
(a) X_1 vs. X_2



(b) X_1 vs. X_3

Scatter Plots for Example 3.2

Figure 2 (continued)



(c) X_2 vs. X_3

Scatter Plots for Example 3.2

We observe that the observations of A are tightly grouped while the remainder are more generally scattered. Thus the self-critical weights were able to separate the observations by a single observer from those of other origins, a feat which the eye would be hard-pressed to match without the aid of the separate labels as given on the scatter plots. Indeed, if the observations were not labeled according to their genesis, one would be tempted to pick out the B's lying at the extremes as being potential problem points and consequently miss the relevant structure entirely. This ability to cluster is also implied by the form of the influence functions which we present in a later section.

4. Some Asymptotics

The estimators $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{V}}$ defined by (2.4)-(2.8) are M-estimators for location and the covariance matrix by virtue of (2.4) and (2.5). Furthermore, if the observations $\underset{\sim}{x}$ are transformed according to $y_i = \underset{\sim}{a} + \underset{\sim}{A}\underset{\sim}{x}_i$, then $\hat{\underset{\sim}{\mu}}_y = \underset{\sim}{a} + \underset{\sim}{A}\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{D}}_y = \underset{\sim}{A}\hat{\underset{\sim}{D}}\underset{\sim}{A}^T$. The estimator $\hat{\underset{\sim}{D}}$ is positive definite with probability one as soon as $n > p$. However, some numerical difficulties may surface if sufficiently many of the final weights $\hat{v}_{ic}$ are approximately zero so that there are *fewer* than p $\hat{v}_{ic}$ which are larger than zero. In this latter case $\hat{\underset{\sim}{V}}$ will be algorithmically singular.

Standard arguments may be used to show that the estimators $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{V}}$ defined as the consistent zeros of (2.4) and (2.5) are asymptotically Gaussian.

Define

$$\ell_{cj} = \frac{1}{c}\left\{\frac{f^c(x_j|\underset{\sim}{\mu},\underset{\sim}{D})}{[Q(\underset{\sim}{\mu},\underset{\sim}{D};c)]^{c/(1+c)}} - 1\right\} \tag{4.1}$$

Then the jth score function is $\frac{\partial \ell_{cj}}{\partial \theta}$ and $E(\frac{\partial \ell_{cj}}{\partial \theta}) = 0$. The self-critical estimators $(\hat{\mu}, \hat{D})$ are asymptotically Gaussian with mean $\underset{\sim}{0}$ and variance-covariance matrix

$$\underset{\sim}{C}_0 = \underset{\sim}{H}_0^{-1} \underset{\sim}{\Sigma}_0 \underset{\sim}{H}_0^{-1} \tag{4.2}$$

where the θ, θ' element of $\underset{\sim}{H}_0$ and $\underset{\sim}{\Sigma}_0$ are respectively

$$H_{0\theta\theta'} = E\left(\frac{\partial^2 \ell_c}{\partial\theta\partial\theta'}\right) = nE\left(\frac{\partial^2 \ell_{cj}}{\partial\theta\partial\theta'}\right) , \tag{4.3}$$

and

$$\Sigma_{0\theta\theta'} = E\left(\frac{\partial \ell_c}{\partial\theta} \frac{\partial \ell_c}{\partial\theta'}\right) = nE\left(\frac{\partial \ell_{cj}}{\partial\theta} \frac{\partial \ell_{cj}}{\partial\theta'}\right) , \tag{4.4}$$

for any $j = 1,2,\ldots,n$ and θ, θ' ranges over $\mu_1, \mu_2, \ldots, \mu_p$ and d_{11}, d_{12}, $\ldots, d_{1p}, d_{22}, \ldots d_{2p}, \ldots, d_{pp}$. The expressions (4.2)-(4.4) are evaluated at the (unknown) true values of $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$, $\underset{\sim}{\mu}_0$, and $\underset{\sim}{D}_0$. When $c = 0$ in (4.2)-(4.4) the expressions are evaluated by a limiting argument and all reduce to the usual maximum likelihood expressions.

The covariance matrix $\underset{\sim}{C}_0$ may be written as

$$\underset{\sim}{C}_0 = \begin{bmatrix} \underset{\sim}{A} & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{B} \end{bmatrix} \tag{4.5}$$

where $\underset{\sim}{A}$ is the p×p covariance matrix for the estimator $\hat{\underset{\sim}{\mu}}$ and $\underset{\sim}{B}$ is the $\frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1)$ covariance matrix for the estimator $\hat{\underset{\sim}{D}}$. The off-diagonal blocks of $\underset{\sim}{0}$ indicate that the estimators $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{D}}$ are asymtotically independent. The matrix $\underset{\sim}{A}$ may be easily evaluated at the multivariate normal distribution and is given by

$$\underset{\sim}{A} = (1+c)^{p+2}(1+c)^{\frac{1}{2}(p+2)}D \tag{4.6}$$

with elements $a_{ij} = \mathrm{cov}(\hat{\mu}_i, \hat{\mu}_j)$. The form of $\underset{\sim}{B}$ is considerably more complex and we evaluate it only for particular cases.

We now give the general form for the matrix $\underset{\sim}{B}$ at the multivariate normal distribution with mean vector $\underset{\sim}{0}$, covariance matrix $\underset{\sim}{I}$. The particular variances and covariances will be indicated by the subscripts used. We have

$$
\begin{aligned}
B_{\hat{d}^2_{jj}} &= (3c^2+4c+2)(c+1)^{p+2}(2c+1)^{-\frac{1}{2}(p+4)}, && \text{for } j=1,\ldots,p; \\
B_{\hat{d}^2_{jk}} &= (c+1)^{p+4}(2c+1)^{-\frac{1}{2}(p+4)} && \text{, for } j \neq k; \\
B_{\hat{d}_{jj}\hat{d}_{kk}} &= c^2(c+1)^{p+2}(2c+1)^{-\frac{1}{2}(p+4)} && \text{, for } j \neq k; \\
B_{\hat{d}_{ij}\hat{d}_{k\ell}} &= 0 && \text{, for all other cases.}
\end{aligned}
\tag{4.7}
$$

Let us consider now the vector of estimators of covariances $\hat{d}_{11}$, $\hat{d}_{12}$, $\hat{d}_{22}$. At the bivariate standard normal with correlation ρ, the matrix $\underset{\sim}{B}$ of (4.5) is given by (4.2) with

$$
H_0 = K_1 \begin{pmatrix} -1 & 2\rho & -\rho^2 \\ 2\rho & -2-2\rho^2 & 2\rho \\ -\rho^2 & 2\rho & -1 \end{pmatrix} \tag{4.8}
$$

and

$$
\underset{\sim}{\Sigma}_0 = K_2 \begin{pmatrix} 3c^3+4c+2 & -2\rho(3c^2+4c+2) & c^2+2\rho^2(c+1)^2 \\ -2\rho(3c^2+4c+2) & 4(c+1)^2+4\rho^2(2c^2+2c+1) & -2\rho(3c^2+4c+2) \\ c^2+2\rho^2(c+1)^2 & -2\rho(3c^2+4c+2) & 3c^2+4c+2 \end{pmatrix} \tag{4.9}
$$

where $K_1 = \frac{1}{2}(2\pi)^{-c}|\underset{\sim}{D}|^{-\frac{1}{2}c-2}(c+1)^{-2}$ and $K_2 = \frac{1}{4}(2\pi)^{-2c}|\underset{\sim}{D}|^{-c-2}(2c+1)^{-3}$.

The general form of (4.8) and (4.9) is quite complicated and is not given here. It can be approximated in practice from the asymptotic results, namely for (4.2)-(4.4) evaluated at $\hat{\underset{\sim}{\mu}}$, $\hat{\underset{\sim}{V}}$.

Table 3 provides the asymptotic efficiencies of the self-critical estimators (c>0) relative to maximum likelihood estimators (c=0) for the estimators $\hat{\mu}_i$, and $\hat{d}_{ij}$ for selected values of the constant c. For fixed c, there is a decrease in the asymptotic relative efficiency for the estimator $\hat{\rho}$ as the correlation ρ increases. The decrease becomes more pronounced for larger values of c. We will see in the sequel that the relative efficiencies of the self-critical estimators approach their asymptotic relative efficiencies rapidly. For samples as small as n = 20, the relative efficiencies are about equal to those given in Table 3.

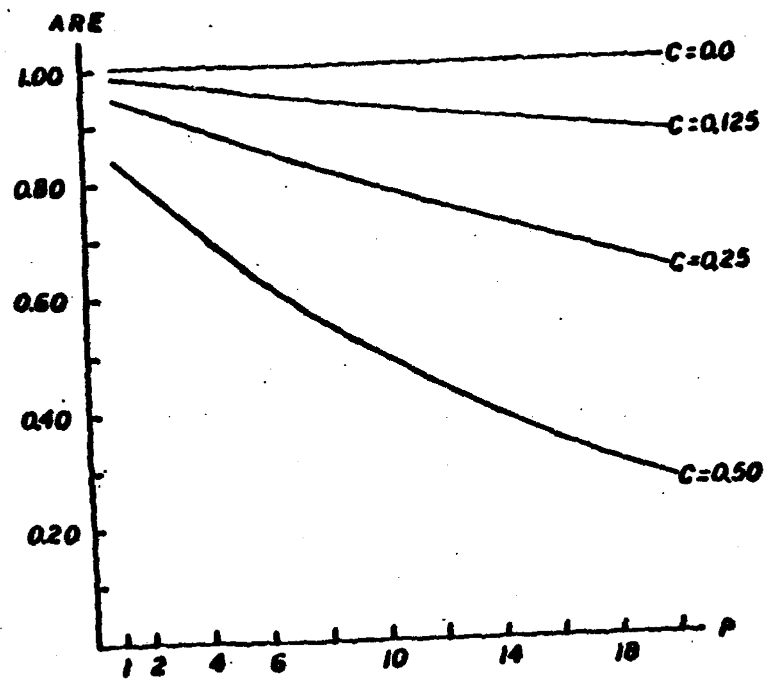
Figures 3a, b, c provide the asymptotic relative efficiences of the self-critical estimators $\hat{\mu}_i$, $\hat{d}_{ii}$, $\hat{\rho}$ as a function of c and dimensionality p when sampling is assumed to be from the p-variate circular normal distribution. For fixed c we see that the efficiency decreases as p increases and that efficiency decreases as c increases. For a given value of c, variance and correlation estimators are less efficient than those of location. It should be emphasized that these figures pertain to the p-dimensional normal with covariance matrix $\underset{\sim}{I}$ only. The efficiencies for covariance estimators will be somewhat lower in the general case as is indicated in Table 3. Maronna (1976) reports asymptotic relative efficiencies for the multivariate version of Huber's proposal 2 at the unit spherical normal for Winsorization proportions $q = .2, .3, .5, 1.0$. It is interesting that, for fixed q, the efficiencies increase with increasing dimensionality.

Table 3

Asymptotic Relative Efficiency (ARE) of the Self-Critical Estimators
for the Bivariate Standard Normal Distribution as a Function of the Index c

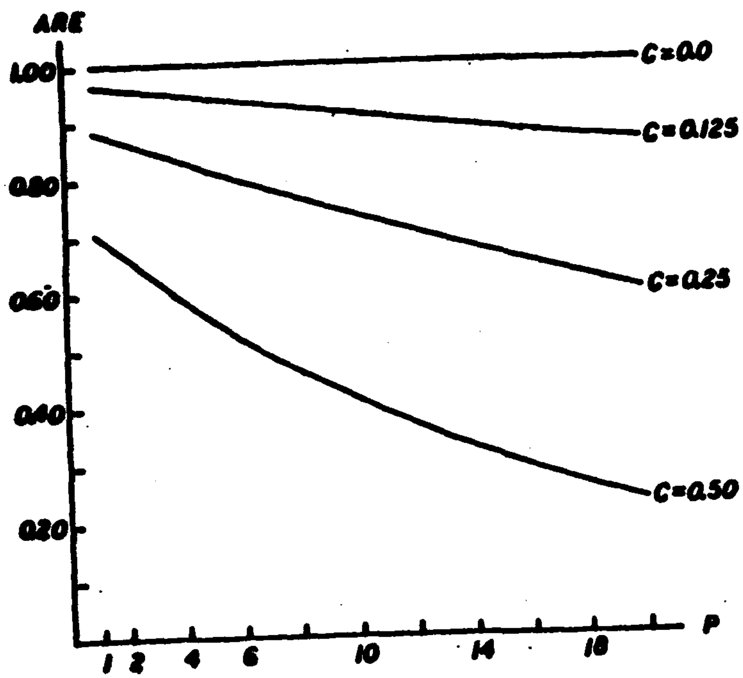
ARE		c												
		0.0	0.1	0.125	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\hat{\mu}_i$		1.0	.983	.975	.945	.922	.896	.843	.790	.739	.690	.644	.602	.563
$\hat{d}_{ii}$		1.0	.971	.958	.906	.867	.827	.744	.665	.593	.528	.470	.420	.375
$\hat{\rho}$	0.0	1.0	.975	.963	.919	.885	.849	.775	.702	.635	.573	.517	.467	.422
	0.1	1.0	.975	.963	.919	.884	.848	.774	.702	.634	.572	.516	.466	.421
	0.3	1.0	.975	.962	.917	.882	.845	.769	.696	.627	.565	.508	.458	.413
	0.5	1.0	.974	.961	.914	.878	.840	.762	.687	.617	.554	.497	.447	.402
	0.7	1.0	.973	.960	.910	.873	.834	.754	.678	.607	.542	.485	.435	.390
	0.9	1.0	.972	.958	.908	.869	.829	.747	.669	.597	.532	.475	.424	.379

Figure 3a.



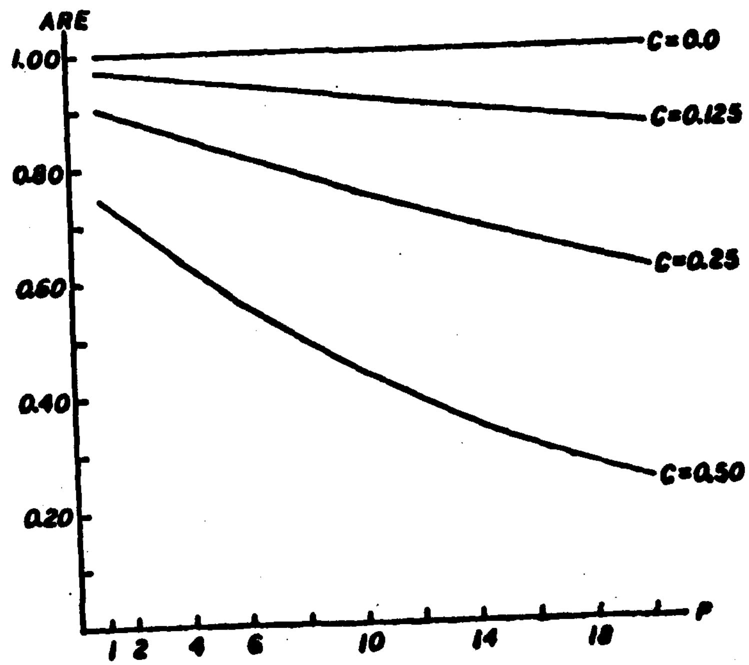
Asymptotic Relative Efficiency of the self-critical location estimate, $\hat{\mu}_i$

Figure 3b.



Asymptotic Relative Efficiency of the self-critical variance estimate $\hat{d}_{11}$.

Figure 3c.



Asymptotic Relative Effficiency of the self-critical correlation estimate $\hat{\rho}$.

5. Influence Functions

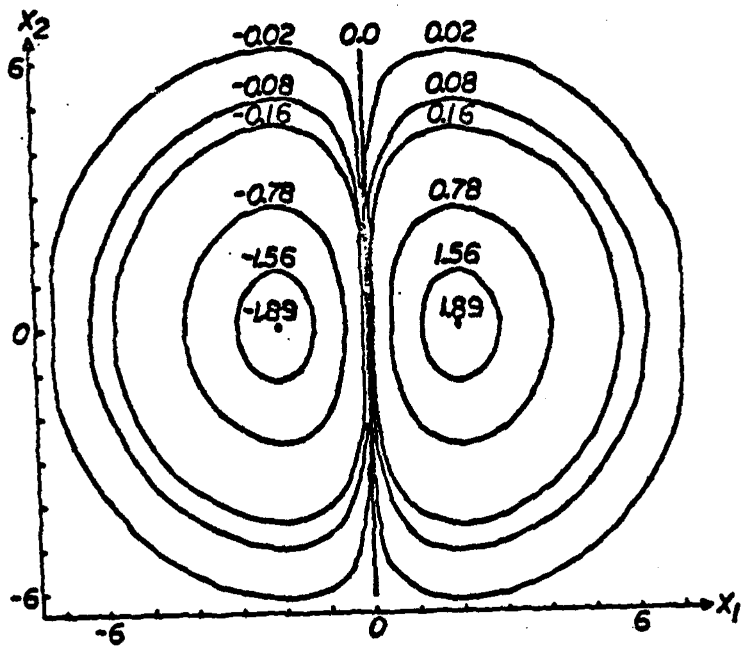
The influence function is the most important determinant of qualitative robustness since many other robustness characteristics of an estimator may be derived from it. The influence function characterizes the (asymptotic) response of an estimator to an additional observation as a function of where the observation falls. The shape of the influence function is determined primarily by the score function $\frac{\partial \ell_{c_j}}{\partial \underset{\sim}{\theta}}$ for $\underset{\sim}{\theta} = \underset{\sim}{\mu}$ or $\underset{\sim}{D}$. In order to guarantee that single (or multiple) observations do not exert an excessive influence on the response of the estimator to single observations, we require that the influence function be bounded. It is also advantageous, in our opinion, if the influence functions are re-descending to zero.

The influence function for $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{D}}$ are both bounded and re-descending to zero for $c > 0$. Let $\ell_{c\underset{\sim}{x}}$ denote the function when argument $\underset{\sim}{x}_j$ of (4.1) is replaced by the non-random vector $\underset{\sim}{x}$. The influence function for $\hat{\underset{\sim}{\mu}}$ at the p-dimensional normal distribution with parameters $\underset{\sim}{\mu}$ and $\underset{\sim}{D}$ is given by

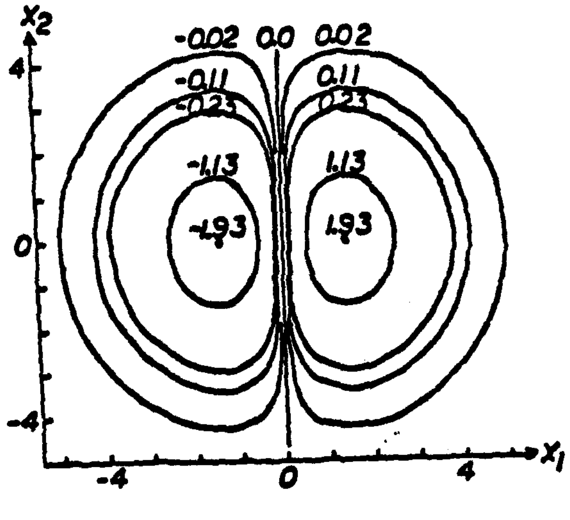
$$\begin{aligned} IC\left(\hat{\underset{\sim}{\mu}}, \underset{\sim}{x}, N_p(\underset{\sim}{\mu}, \underset{\sim}{D})\right) &= - E\left(\frac{\partial^2 \ell_{c_j}}{\partial \underset{\sim}{\mu} \partial \underset{\sim}{\mu}^T}\right)^{-1} \frac{\partial \ell_{c_j}}{\partial \underset{\sim}{\mu}} \\ &= (1+c)^{\frac{1}{2}(p+2)} (\underset{\sim}{x}-\underset{\sim}{\mu}) \exp\left(-\frac{c}{2} (\underset{\sim}{x}-\underset{\sim}{\mu})^T \underset{\sim}{V}^{-1} (\underset{\sim}{x}-\underset{\sim}{\mu})\right) , \end{aligned} \tag{5.1}$$

for any $j = 1,2,\ldots,n$. This influence function is illustrated in Figures 4a and b where we exhibit contours of the first component of the vector (5.1) with $c = \frac{1}{4}$ and $c = \frac{1}{2}$ respectively at the standard bivariate normal distribution with zero correlation. These figures and other figures not given here indicate that to choose a greater degree of robustness we choose a larger value of c. The contour

Figure 4



(a) c = ¼



(b) c = ½

Influence curve contours for the self-critical location estimate, $\hat{\mu}_1$, at the bivariate standard normal distribution with correlation $\rho = 0$

for the first component of the self-critical estimator $\hat{\underset{\sim}{\mu}}$ for the standard bivariate normal with $\rho = .9$ and $c = ¼$ is shown in Figure 5. The nearly elliptical contours have the line of perfect positive correlation as their major axis. Moderately distant observations that are close to this axis have greater influence on the location estimator than those that are closer but along the minor axis. Note also that the component x_2 of $\underset{\sim}{x}$ has an influence on the estimator $\hat{\mu}_1$, the first component of $\underset{\sim}{\mu}$.

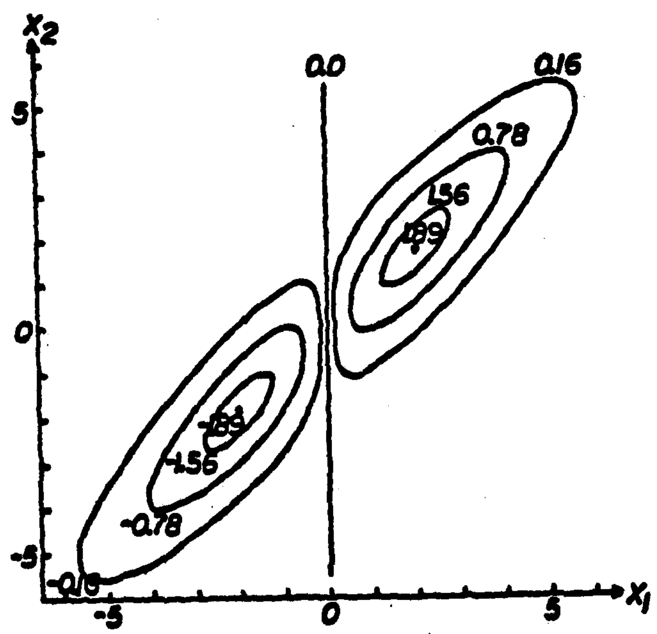
The influence function for the self-critical estimator $\hat{\underset{\sim}{D}}$ cannot be written in simple form in general. For the p = 2 case we illustrate it in terms of the vector of estimators $\hat{\underset{\sim}{w}}^T = (\hat{v}_{11}, \hat{v}_{12}, \hat{v}_{22})$ for the standard bivariate normal distribution with correlation ρ. It may be shown that

$$IC(\hat{\underset{\sim}{w}}, \underset{\sim}{x}, N_2(0, \underset{\sim}{P}) = (1+c)^2 \begin{pmatrix} (1+c)x_1{}^2 - 1 \\ (1+c)x_1x_2 - \rho \\ (1+c)x^2 - 1 \end{pmatrix} \exp(-\frac{c}{2} \underset{\sim}{x}^T \underset{\sim}{P}^{-1} \underset{\sim}{x})$$

where $\underset{\sim}{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. We present contours of this function for $\hat{d}_{ij}$ with $c = ¼$ for correlations $\rho = 0$ and .9 in Figures 6a and b. In Figures 7a and b we show contours of the influence function for the self-critical estimator of ρ under the same conditions.

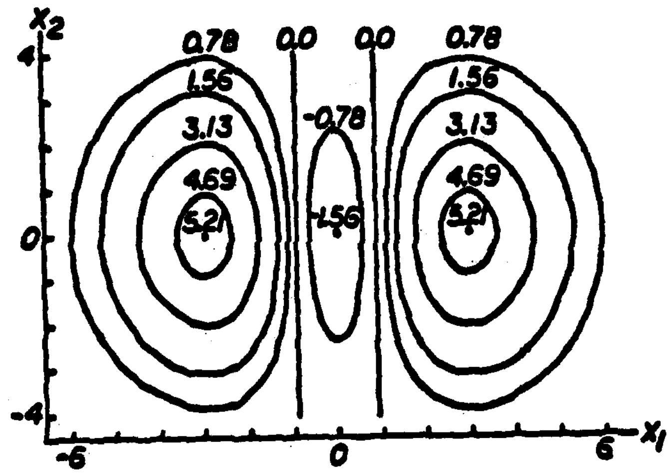
All the influence functions are bounded for $c > 0$ and are characterized by closed contours. The closed contours indicate that the self-critical procedure will be useful in clustering data. The clustering nature of the estimation procedures explains why we were able to identify the 13 observations by the same individual in example 3.2. Just as in this example, the procedure will generally focus on the tightest cluster of observations as the parameter c is increased. The estimation procedure

Figure 5

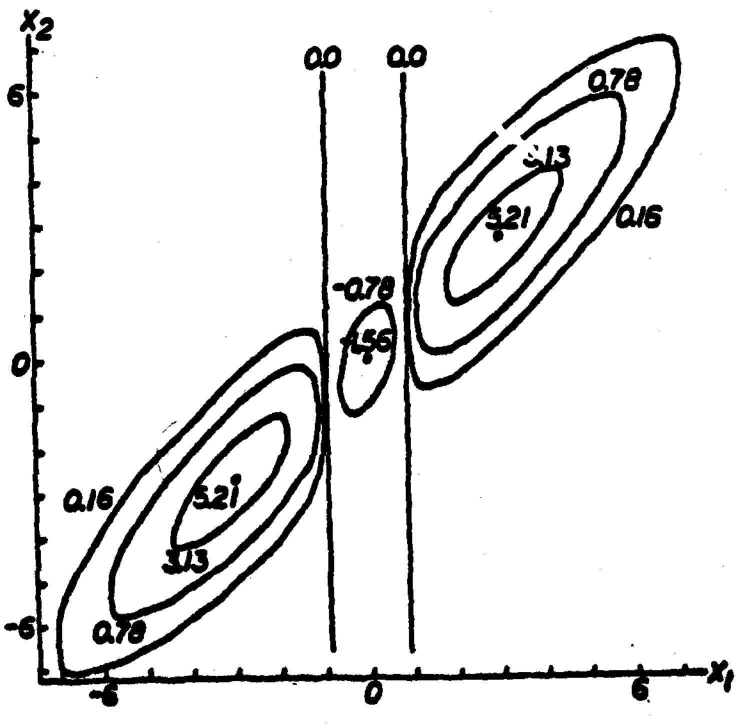


Influence curve contours for the self-critical location estimate, $\hat{\mu}_1$, at the bivariate standard normal distribution with correlation $\rho = 0.9$, $c = \frac{1}{4}$

Figure 6



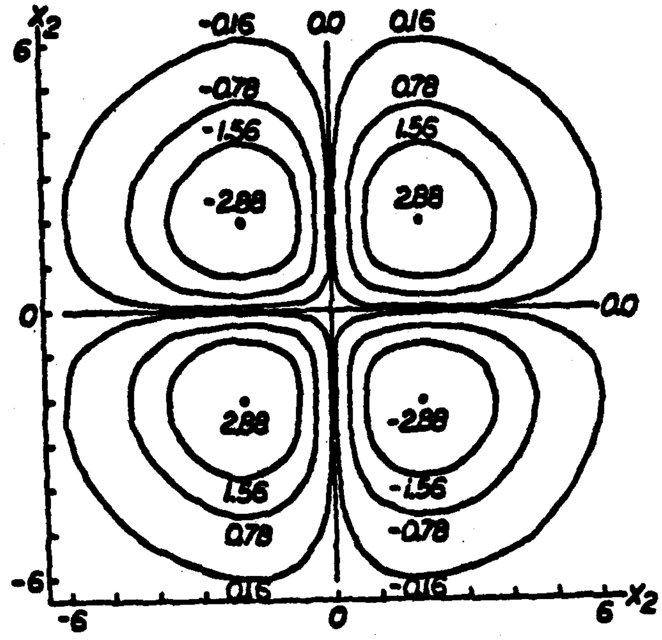
(a) $\rho = 0.0$



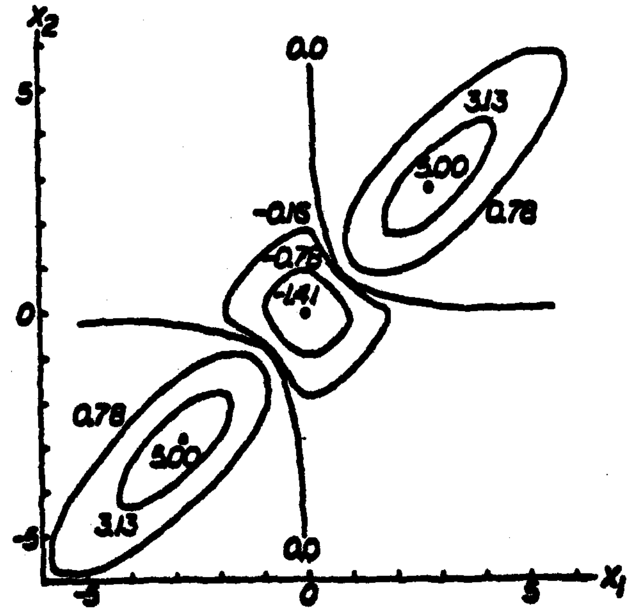
(b) $\rho = 0.9$

Influence curve contours for the self-critical variance estimate, $\hat{d}_{11}$, at the bivariate standard normal distribution, $c = \frac{1}{4}$

Figure 7



(a) $\rho = 0.0$



(b) $\rho = 0.9$

Influence curve contours for the self-critical correlation estimate, $\hat{\rho}$, at the bivariate standard normal distribution, $c = \frac{1}{2}$

may also be useful in classification problems, but we do not pursue the subject here.

Devlin et al. (1975) present the influence functions for the maximum likelihood estimator of ρ and a 5% trimmed estimator. Neither of these influence functions are bounded, and the trimmed estimator is unable to limit the influence of observations in certain regions of the $\underset{\sim}{x}$ space.

6. Monte Carlo Study

It has been long recognized that some alternative univariate estimators such as the α-trimmed mean lose little efficiency for an exactly normal parent but exhibit much better and more stable performance under departures from it (Huber, 1972, for example). To evaluate the performance of the estimators of section 2 with $c > 0$, a limited Monte Carlo study was conducted. Simulated bivariate normal samples were used to gauge losses in efficiency and slightly non-normal samples were used to measure stability and performance. The performances of the usual vector of sample means and covariance matrix, the vector of 10% (on each side) trimmed means combined with a covariance matrix estimator employing 5% (on each side) trimmed variances and the estimators $\hat{\underset{\sim}{\mu}}$ and $\hat{\underset{\sim}{V}}$ for several values of c are examined under several sampling situations.

To study the performance of the selected estimators in sampling from normal populations, 100 samples of size 20 from each of the bivariate normal populations with $\mu_2 - \mu_1 = 0$, $\sigma_1^2/\sigma_2^2 = 0.5, 1.0, 3.0$ and $\rho = 0.0, 0.5, 0.9, 0.99$ were constructed. The same random number seed was used to generate samples for all correlations and variance ratios.

Tables of the Monte Carlo means and variances of the population parameters estimated are presented in Table 4. In this and the other tables of this section MEAN represents the sample mean vector or the usual sample variances and correlations, TR. MEAN represents the 10% on each side) vector of trimmed means or the 5% trimmed variances and correlations and SC($\frac{1}{8}$), SC($\frac{1}{4}$), SC($\frac{1}{2}$) represent the self-critical estimators of section 2 with parameters $c = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}$ respectively. All estimators provide reasonable estimates for location, but SC($\frac{1}{2}$) and TR. Mean have slightly inferior values vis-à-vis the other estimators. The $c = 0$ and $c = \frac{1}{8}$ cases have the smallest variances. Comparisons of the Monte Carlo means and variances for the variance ratio and correlation yield similar conclusions. Note that the efficiencies computed by dividing the $c = 0$ variances by the $c > 0$ variances are generally good approximations to the values given in Table 3. An exception occurs for $c = \frac{1}{2}$ for the estimator $\hat{\rho}$ of correlation. An analogous characterization of the estimators is drawn if mean squared errors are used instead of variances, but we do not present the results here.

To study stability of the estimators under slight departures from a Gaussian parent, bivariate normal samples were constructed and contaminants of one of two types were added to the basic normal variates. In the first case, one simulated bivariate Cauchy was added to standard bivariate Gaussian samples of size 19. In the second case, a mixture of a random number r of bivariate Gaussian variates divided by a uniform deviate

TABLE 4

Monte Carlo Results for Bivariate Normal Samples of Size 20

		Monte Carlo Means					Monte Carlo Variances				
σ_1^2/σ_2^2	ρ	MEAN	TR.MEAN	SC(1/2)	SC(1/4)	SC(1/8)	MEAN	TR.MEAN	SC(1/2)	SC(1/4)	SC(1/8)
<u>(a) Location</u>											
0.5		-.022	-.011	-.006	-.017	-.020	.102	.118	.127	.107	.102
1.0		-.016	-.007	-.001	-.012	-.014	.056	.063	.069	.059	.056
3.0		-.017	-.006	-.003	-.012	-.015	.076	.087	.094	.080	.076
<u>(b) Variance</u>											
0.5		.513	.501	.508	.510	.512	.357	.418	.527	.382	.342
1.0		1.027	1.012	1.025	1.020	1.024	.103	.122	.143	.109	.098
3.0		3.081	3.037	3.047	3.062	3.073	.254	.307	.367	.250	.231
<u>(c) Correlation</u>											
	0.00	-.014	-.006	.001	-.021	-.022	.055	.056	.098	.069	.062
	0.50	.480	.481	.477	.480	.482	.040	.038	.070	.043	.040
	0.90	.895	.892	.885	.892	.895	.003	.004	.013	.003	.003
	0.99	.990	.989	.989	.989	.990	.4(-4)	.5(-4)	.8(-4)	.4(-4)	.4(-4)

and n - r bivariate Gaussian variate was constructed. The number r was generated from a binomial distribution with parameters n = 20 and probability .05. The parameters of the n - r bivariate Gaussian variates were $\mu_2 - \mu_1 = 0$, $\sigma_1^2/\sigma_2^2 = 1$ and 3, and $\rho = 0, .5, .9, .99$. Table 5 depicts the Monte Carlo means and variances for each estimator, based on 25 samples generated from each population. The Monte Carlo means for location, variance, and correlation show that the self-critical estimators and the trimmed estimators retain their integrity with respect to the basic underlying assumption of Gaussianity. This is, however, not as important as the fact that this stability concerning the basic cluster allows potential contaminants or model departures of various types to be identified.

A range of experience with both real and simulated data corroborates the above results. The estimators perform well in practice, especially when used in the context of a sensitivity analysis.

TABLE 5

Monte Carlo Results for Non-Normal Samples of Size 20

	$\sigma_1^2\sigma_1^2$	ρ	Monte Carlo Means MEAN	TR.MEAN	SC(1/2)	SC(1/4)	SC(1/8)	Monte Carlo Variances MEAN	TR.MEAN	SC(1/2)	SC(1/4)	SC(1/8)
(a) Location												
B.C.	1.0		0.100	0.070	0.032	0.017	0.018	0.149	0.055	0.068	0.051	0.04
N/U	1.0		0.258	-0.105	0.025	0.004	0.001	4.412	0.068	0.086	0.064	0.064
N/U	3.0		0.251	0.048	-0.009	-0.041	-0.058	0.590	0.144	0.199	0.145	0.138
(b) Variance												
B.C.	1.0		1.533	0.974	0.935	0.961	0.980	12.72	0.130	0.157	0.118	0.135
N/U	1.0		0.040	6.666	1.063	1.091	1.133	5.(+6)	2.(+2)	0.173	0.121	0.174
N/U	3.0		0.683	3.075	2.742	2.887	2.990	1.(+4)	1.316	1.067	1.014	1.089
(c) Correlation												
		0.00	0.012	-0.022	0.008	-0.005	-0.007	0.162	0.054	0.096	0.061	0.057
B.C.	1.0	0.50	0.540	0.449	0.463	0.473	0.477	0.061	0.026	0.051	0.032	0.034
		0.90	0.906	0.883	0.886	0.892	0.892	0.006	0.002	0.003	0.002	0.002
		0.99	0.990	0.988	0.989	0.989	0.989	0.1(-3)	1.2(-4)	0.3(-4)	0.2(-4)	0.3(-4)
		0.00	0.016	0.068	-0.001	0.029	0.023	0.093	0.060	0.114	0.046	0.049
N/U	1.0	0.50	0.469	0.555	0.518	0.536	0.540	0.057	0.030	0.065	0.024	0.024
		0.90	0.799	0.898	0.894	0.907	0.904	0.051	0.004	0.010	0.002	0.003
		0.99	0.876	0.978	0.990	0.991	0.990	0.049	0.002	0.1(-3)	0.3(-4)	0.3(-4)
		0.00	-0.046	-0.010	-0.036	-0.029	-0.030	0.053	0.081	0.125	0.055	0.044
N/U	3.0	0.50	0.414	0.454	0.413	0.441	0.444	0.055	0.062	0.078	0.046	0.032
		0.90	0.819	0.873	0.874	0.879	0.881	0.043	0.008	0.005	0.004	0.004
		0.99	0.923	0.983	0.987	0.987	0.987	0.039	0.001	0.5(-4)	0.5(-4)	0.5(-4)

7. Multivariate Regression

We now show that structured multivariate data, specifically a linear model, may be naturally treated by the self-critical approach of section 2. Let $\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots, \underset{\sim}{y}_p$ be N×1 vectors representing N independent observations on each of p correlated random variables. We assume the model representation

$$\underset{\sim}{y}_j = \underset{\sim}{X}\underset{\sim}{\beta}_j + \underset{\sim}{U}_j \ , \ j = 1,2,\ldots, p \tag{7.1}$$

where $\underset{\sim}{X}$ is N×q and considered given, $\underset{\sim}{\beta}_j$ is q×1 and unknown, and $\underset{\sim}{U}_j$ is an error or disturbance vector which is distributed as $N_p(\underset{\sim}{0},\underset{\sim}{D})$. For a single value of j, (7.1) is a univariate regression model. Define

$$\underset{(N\times p)}{\underset{\sim}{Y}} = (\underset{\sim}{y}_1, \underset{\sim}{y}_2, \ldots, \underset{\sim}{y}_p),\ \underset{(N\times q)}{\underset{\sim}{U}} = (\underset{\sim}{u}_1, \underset{\sim}{u}_2, \ldots, \underset{\sim}{u}_p),\ \underset{(q\times p)}{\underset{\sim}{B}} = (\underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \ldots \underset{\sim}{\beta}_p)$$

and then (7.1) may be written as

$$\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{B} + \underset{\sim}{U} \ . \tag{7.2}$$

We wish to estimate $\underset{\sim}{D}$ and $\underset{\sim}{B}$. Let $\underset{\sim}{y}_i$, $\underset{\sim}{x}_i$, and $\underset{\sim}{u}_i$ denote the row vectors of $\underset{\sim}{Y}$, $\underset{\sim}{X}$, and $\underset{\sim}{U}$ of dimensions 1×p, 1×q, and 1×p respectively. Since $\underset{\sim}{u}_i = \underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B}$, the density of the $\underset{\sim}{u}_i$ is given by

$$f(\underset{\sim}{u}_i) \equiv f(\underset{\sim}{u}_i|\underset{\sim}{0},\underset{\sim}{D}) = |2\pi\underset{\sim}{D}|^{-\frac{1}{2}} \exp\left(-\tfrac{1}{2}(\underset{\sim}{y}_i-\underset{\sim}{x}_i\underset{\sim}{B})\underset{\sim}{D}^{-1}(\underset{\sim}{y}_i-\underset{\sim}{x}_i\underset{\sim}{B})^T\right) \tag{7.3}$$

The objective function to be maximized for $\underset{\sim}{B}$ and $\underset{\sim}{D}$ is, from section 2,

$$\ell_c = \frac{1}{c}\sum_{i=1}^{N}\left[\frac{f^c(\underset{\sim}{u}_i)}{\{Q(\underset{\sim}{B},\underset{\sim}{D};c)\}^{c/(1+c)}} - 1\right] \ . \tag{7.4}$$

On differentiating with respect to B and D, setting the resulting expressions to 0, and solving we find that B and D satisfy the implicit equations

$$\underset{\sim}{B} = \left\{ \sum_{i=1}^{N} \underset{\sim}{x}_i^T \underset{\sim}{x}_i \exp\left(-\frac{c}{2} (\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B}) \underset{\sim}{D}^{-1} (\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B})^T \right) \right\}^{-1} \times$$

$$\left\{ \sum_{i=1}^{N} \underset{\sim}{x}_i^T \underset{\sim}{y}_i \exp\left(-\frac{c}{2} (\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B}) \underset{\sim}{D}^{-1} (\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B})^T \right) \right\} \tag{7.5}$$

and

$$\underset{\sim}{D} = (1+c) \sum_{i=1}^{N} w_i (\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B})^T (\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B}) \quad , \tag{7.6}$$

where

$$w_i = \frac{f^c(\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B};\ \underset{\sim}{0}, \underset{\sim}{D})}{\sum_{i=1}^{N} f^c(\underset{\sim}{y}_i - \underset{\sim}{x}_i \underset{\sim}{B};\ \underset{\sim}{0}, \underset{\sim}{D})} \quad , \ i=1,2,\ldots,N. \tag{7.7}$$

It is clear that if $c = 0$, these equations reduce to the usual maximum likelihood equations. For $c > 0$, the estimators are not equation-by-equation univariate estimators since the covariance structure of $\underset{\sim}{D}$ is now allowed to play a role in the estimation process. It is interesting that, from (7.5) to (7.7),

$$\begin{aligned} \sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \hat{\underset{\sim}{u}}_i &= \sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \underset{\sim}{y}_i - \sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \underset{\sim}{x}_i \ \hat{\underset{\sim}{B}} \\ &= \sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \underset{\sim}{y}_i - \sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \underset{\sim}{x}_i \left(\sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \underset{\sim}{x}_i \right)^{-1} \sum_{i=1}^{N} \underset{\sim}{w}_i \underset{\sim}{x}_i^T \underset{\sim}{y}_i = \underset{\sim}{0}. \end{aligned} \tag{7.8}$$

In the same manner we have

$$\sum_{i=1}^{N} \underset{\sim}{w}_i (\underset{\sim}{x}_i \hat{\underset{\sim}{B}})^T \hat{\underset{\sim}{u}}_i = \underset{\sim}{0} \quad . \tag{7.9}$$

The equations (7.8) and (7.9) are weighted versions of the usual orthogonality relations associated with the maximum likelihood or least squares procedures.

This regression procedure with $c > 0$ is useful in providing protection against difficulties in the $\underset{\sim}{y}$ values, for example $\underset{\sim}{y}$ outliers, but is not especially useful if there are difficulties associated with the factor space, for example, $\underset{\sim}{x}$ outliers. This is easily seen by rewriting (7.5) in terms of the score function for $\underset{\sim}{B}$. This score function is bounded and re-descending in the residuals $\underset{\sim}{y}_i - \underset{\sim}{x}_i\underset{\sim}{B}$ but is unbounded in the $\underset{\sim}{x}_i$. Thus a single bad $\underset{\sim}{x}_i$ value may ruin an entire regression analysis. An extension of the approach we provide here will be useful in dealing with this problem, but we do not pursue it here. Belsley, Kuh and Welsch (1980) provide a nice discussion of the univariate version of this problem.

For completeness, we provide an example in which we compare the regression equations obtained by maximum likelihood and those obtained from (7.5)-(7.7). It is found that the solutions are sensitive to the change in c from 0 to $\frac{1}{4}$ and thus there are some potential difficulties with the data or the model or both. It is our job, having been forewarned, to find out where the difficulties might be. However, this is not always possible.

<u>Example 7.1</u>: Data extracted from a study of the effects of a change in environment on blood pressure is included in Ryan et al. (1976). In this study, anthropologists measured the blood pressure and other characteristics of Peruvian Indian males over age 21, who had migrated

from primitive environments at high altitudes to modern lower altitude areas of Peru. Previous studies in Africa suggested that such migrations might increase blood pressure at first, but that the blood pressure would tend to return to normal with time. The 39 observations included in this reference are listed in Table 6; the dependent variables are systolic and diastolic blood pressure and the independent variables chosen from those available are F, the fraction of life in the new environment, W, the weight in kilograms, and S, a skin-fold measure of general obesity.

We will consider the linear model containing the aforementioned independent variables and a constant term. The MLE for the full model are

$$\hat{\underset{\sim}{B}} = \begin{bmatrix} 55.51 & 33.21 \\ -28.41 & -9.26 \\ 1.39 & 0.74 \\ -0.26 & 0.001 \end{bmatrix}, \qquad \hat{\underset{\sim}{\Sigma}} = \begin{bmatrix} 85.62 & 28.58 \\ 28.58 & 108.36 \end{bmatrix}.$$

The order in the matrix of parameter estimates is constant, fraction of life, weight, and obesity measure with the first column representing the systolic blood pressure equation and the second column representing the diastolic blood pressure equation. The self-critical estimates with parameter $c = \frac{1}{4}$ are

$$\hat{\underset{\sim}{B}} = \begin{bmatrix} 58.59 & 34.91 \\ -21.31 & -1.08 \\ 1.28 & 0.58 \\ -0.24 & 0.22 \end{bmatrix}, \qquad \hat{\underset{\sim}{\Sigma}} = \begin{bmatrix} 76.98 & 17.26 \\ 17.26 & 65.57 \end{bmatrix}.$$

Table 6

Characteristics of Peruvian Indians in Example 7.1

Observation Number	Blood Pressure Sys.	Blood Pressure Dia.	F	W	S	SC Wt.	RESIDUALS c = 0 $Y_{1_j}-\hat{Y}_{1_j}$	RESIDUALS c = 0 $Y_{2_j}-\hat{Y}_{2_j}$	RESIDUALS c = 1/4 $Y_{1_j}-\hat{Y}_{1_j}$	RESIDUALS c = 1/4 $Y_{2_j}-\hat{Y}_{2_j}$
1	170	76	.048	71.0	27.7	.007	24.25	- 9.10	28.04	- 6.13
2	120	60	.273	56.5	16.3	.026	- 2.12	-12.32	- 1.33	-10.84
3	125	75	.208	56.0	8.9	.031	- 0.14	2.46	1.13	5.94
4	148	120	.042	61.0	11.0	.4(-3)	11.71	42.23	14.67	47.40
5	140	78	.040	65.0	42.4	.027	6.13	- 2.77	9.18	- 3.96
6	106	72	.704	62.0	12.0	.023	-12.62	- 0.38	-14.26	- 0.69
7	12[illegible]	76	.179	53.0	20.0	.031	1.02	5.38	2.06	6.15
8	108	62	.893	53.0	8.0	.031	6.24	- 1.99	2.35	- 4.39
9	124	70	.194	65.0	29.3	.027	- 8.86	- 9.34	- 6.75	- 8.86
10	134	64	.406	57.0	15.7	.019	14.83	- 7.45	14.73	- 6.95
11	116	76	.394	66.5	20.0	.023	-15.63	- 2.57	-14.68	- 1.42
12	114	74	.303	59.1	22.3	.029	- 9.34	0.03	- 8.56	0.24
13	130	80	.441	64.0	15.6	.031	2.07	3.71	2.46	5.06
14	118	68	.514	69.5	21.3	.021	-14.04	-11.67	-13.65	-11.32
15	138	78	.057	64.0	13.4	.032	- 1.41	- 1.84	1.74	3.14
16	134	86	.333	56.5	20.0	.018	14.55	14.23	14.86	14.29
17	120	70	.417	57.0	12.0	.033	0.18	- 1.35	0.04	- 0.11
18	120	76	.432	55.0	16.3	.030	4.50	6.26	4.00	6.10
19	114	80	.460	57.0	23.0	.029	- 1.79	9.03	- 2.35	7.47
20	124	64	.263	58.0	27.7	.025	2.44	- 9.53	3.[illegible]3	-10.38
21	114	66	.474	59.5	17.0	.029	- 6.40	- 6.67	- 6.73	- 6.62
22	136	78	.290	61.0	11.3	.029	6.83	2.52	8.03	5.60
23	126	72	.290	57.0	9.0	.032	1.80	- 0.53	2.60	2.43
24	124	62	.539	57.5	12.0	.026	6.94	- 8.59	6.00	- 8.27
25	128	84	.615	74.0	29.3	.031	- 5.38	1.94	- 5.32	0.39
26	134	92	.359	72.0	27.3	.026	- 4.39	9.04	- 2.70	9.72
27	112	80	.610	62.5	19.3	.026	- 8.11	6.38	- 9.12	5.28
28	128	82	.781	68.0	26.3	.031	6.88	5.90	5.17	2.71
29	134	92	.122	63.4	23.3	.021	- 0.20	13.19	2.31	15.34
30	128	90	.286	68.0	28.7	.025	- 6.56	9.31	- 4.79	9.64
31	140	72	.581	69.0	14.0	.026	8.69	- 6.67	8.64	- 5.32
32	138	74	.605	73.0	21.7	.028	3.76	- 7.41	3.88	- 7.34
33	118	66	.233	64.0	15.7	.022	-15.83	-12.22	-13.96	- 9.19
34	110	70	.432	65.0	22.4	.020	-17.85	- 7.12	-17.36	- 7.05
35	142	84	.409	71.0	12.0	.030	2.51	2.27	3.91	5.79
36	134	70	.222	60.2	9.3	.031	3.52	- 5.51	5.13	- 1.56
37	116	54	.021	55.0	10.0	.020	-12.79	-19.54	-10.30	-14.93
38	132	90	.860	70.0	21.7	.026	9.19	13.16	7.17	10.67
39	152	88	.741	87.0	34.3	.029	5.39	- 2.48	5.88	- 4.11

The parameter estimates obtained via the two methods are not strikingly different. The only obvious difference is in the estimate for the fraction of life term for both systolic and diastolic blood pressure equations. Table 6 also includes the weight assigned to each observation by the SC procedure. Observations 4 and 1 received the lowest final weights, $\hat{w}_{ic}$. The 22-element of the covariance matrix is quite sensitive to changes in the value of c and the residuals of Table 6 confirm this.

The negative *coefficients* for the fraction of life in the new environment support the results of the African studies, showing that blood pressure decreases with the length of time in the new environment. However, it appears that there is less of a decrease in the diastolic measurement than in the systolic measurement. As expected, weight has the effect of increasing the blood pressure measures, but it appears to have a greater effect on the systolic measurement.

The four term linear model is not entirely appropriate since there are still patterns in both the $c = 0$ and $c = \frac{1}{4}$ residuals. In this case examination of the $c = 0$ residuals would have led to the two potential problem points, 1 and 4, that the final weights $\hat{w}_{i,.25}$ have flagged for further scrutiny, but this will frequently not be the case.

Our procedure may be directly applied to multivariate analysis of variance. A generalized inverse may be used in place of the inverse of (7.5) but a different, more direct, approach seems preferable.

8. Discussion

We have presented a procedure for the analysis of multivariate normal data which is at once very general and easy to use. We have successfully used it in a variety of applications. It has proved to be helpful in the identification of potential outliers and in the process of model evolution. The procedure will tolerate large amounts (more than 100%) of contamination, provided a significant portion of it is not all tightly concentrated in a single area of p-dimensional space. This is of little consequence in any event because the final weights will show a dramatic sensitivity to changes in the index c and the source of the sensitivity should always be sought. The extent to which the index c may differ from zero is a function of the dimension p and the sample size n. For example, if $p = 10$ and $n = 200$, the procedure is likely to break down with $c > .15$. It is highly recommended that the procedure proposed be used in an exploratory fashion.

REFERENCES

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. Wiley, New York.

Barnett, V. and Lewis, T. (1978). Outliers in Statistical Data. Wiley, New York.

Billmeyer, F.W. and Rich, D. (1978). Unpublished data from color matching on a three-primary additive colorimeter. Rensselaer Polytechnic Institute, Dept. of Chemistry, Troy, New York.

Brown, W.R.J. et al. (1956). Multivariate Normality of the Color-Matching Process. J. Optical Society of America 46, 46-49.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1975). Robust Estimation and Outlier Detection with Correlation Coefficients. Biometrika 62, 531-44.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981). Robust Estimation of Dispersion Matrices and Principal Components. J. American Statistical Association 76, 354-62.

Gnanadesikan, R. (1977). Methods for Statistical Analysis of Multivariate Observation. Wiley, New York.

Gnanadesikan, R. and Kettenring, J.R. (1972). Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data. Biometrics 28, 8--124.

Huber, P.J. (1972). Robust Statistics: A Review. Annals of Mathematical Statistics 43, 1041-67.

Huber, P.J. (1977). Robust Statistical Procedures. SIAM, Philadelphia, Pa.

Huber, P.J. (1981). Robust Statistics. Wiley, New York.

Maronna, R.A. (1976). Robust M-estimators of Multivariate Location and Scatter. Annals of Statistics 4, 51-67.

Mosteller, F. and Tukey, J.W. (1977). Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley, Reading, Mass.

Paulson, A.A., Presser, M., and Nicklin, E.H. (1982). "Self-Critical, and Robust, Procedures for the Analysis of Univariate Complete Data." Submitted for publication.

Rao, C.R. (1965). Linear Statistical Inference and Its Applications. Wiley, New York.

Ryan, T.A., et al. (1976). MINITAB Student Handbook. Duxbury Press, Mass.